Description of the MHS Health Level 7 Chemistry Laboratory for Public Health Surveillance

Technical Document NMCPHC-EDC-TD-1-2014

By Gosia Nowak
EpiData Center Department
September 2012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
30-09-2012	Technical Document	May 2004-September 2012

4. TITLE AND SUBTITLE
Description of the MHS Health Level 7 Chemistry Laboratory for Public Health Surveillance

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Gosia Nowak

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy and Marine Corps Public Health Center
EpiData Center Department
620 John Paul Jones Circle, Suite 1100
Portsmouth, VA 23708

8. PERFORMING ORGANIZATION REPORT NUMBER
NMCPHC-EDC-TD-1-2014

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Navy and Marine Corps Public Health Center
EpiData Center Department
620 John Paul Jones Circle, Suite 1100
Portsmouth, VA 23708

10. SPONSOR/MONITOR'S ACRONYM(S)
NMCPHC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NMCPHC-EDC-TD-1-2014

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release. Distribution is not limited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The EpiData Center Department (EDC) at the Navy and Marine Corps Public Health Center(NMCPHC) evaluated the Health Level 7 (HL7) data source for its usefulness in health surveillance activities. This technical document provides a history of the HL7 chemistry database and its contents, explains the creation of chemistry/serology records, describes the pathway of data from healthcare provider to the EDC, provides a detailed descriptions of all variables within the database, and assesses the database's strengths and limitations. Given an understanding of the strengths and limitations of the data, HL7 chemistry laboratory data have proven to be a valuable source of health information for surveillance purposes. The data allow the creation of a timeline of events corresponding to a specific disease occurrence. Furthermore, data are received in a timely fashion, allowing for near-real-time surveillance of diseases.

15. SUBJECT TERMS
Health Level 7 (HL7), Chemistry, Serology, Surveillance

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	41	757-953-0979

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Abstract

The EpiData Center Department (EDC) at the Navy and Marine Corps Public Health Center(NMCPHC) evaluated the Health Level 7 (HL7) data source for its usefulness in health surveillance activities. This technical document provides a history of the HL7 chemistry database and its contents, explains the creation of chemistry/serology records, describes the pathway of data from healthcare provider to the EDC, provides a detailed descriptions of all variables within the database, and assesses the database's strengths and limitations. Given an understanding of the strengths and limitations of the data, HL7 chemistry laboratory data have proven to be a valuable source of health information for surveillance purposes. The data allow the creation of a timeline of events corresponding to a specific disease occurrence. Furthermore, data are received in a timely fashion, allowing for near-real-time surveillance of diseases.

Contents

List of Tables and Figures

Executive Summary

Background
The EpiData Center Department (EDC) at the Navy and Marine Corps Public Health Center (NMCPHC) was tasked by Global Emerging Infections Surveillance and Response System (GEIS) with the evaluation of the Health Level 7 (HL7) data source for its usefulness in health surveillance activities. This technical document discusses the chemistry database by providing a history of the dataset and its contents, explaining the creation of records, describing the pathway from provider to the EDC, and assessing its strengths and limitations.

Department Of Defense (DOD) Public Health Surveillance Applications
The HL7 chemistry database has been used extensively by the EDC for daily case finding of reportable diseases, preparing health reports, and responding to congressional requests for disease burden. Positive results can be matched with ambulatory or inpatient records to identify how specific laboratory tests correlate with encounters. Furthermore, laboratory confirmation along with clinical compatibility from encounters may help with case validation or confirmation.

Key Fields for Public Health Surveillance
Using multiple fields within the HL7 chemistry database assists in queries and surveillance based on disease context. Combinations of sponsor's social security number (SSN), Family Member Prefix (FMP), and/or unique specimen identifiers are used to identify unique cases or individuals. The date and time variables distinguish timeframes between different laboratory events.

Strengths
HL7 chemistry data are valuable as a source of health information for surveillance purposes. The data are organized in a logical and informative manner, the variable fields of highest importance are complete and contain expected values, and data are received by EDC in a timely manner. This database is useful for the following factors: 1) ability to establish a timeline of events based on laboratory results, 2) timeliness of data, 3) efficiency in data structuring, and 4) increased accuracy of data with only completed and certified entries from DHSS.

Limitations
HL7 data are generated within the Composite Health Care System (CHCS) at fixed-military treatment facilities (MTFs). A typical analysis would not include chemistry records from purchased care, shipboard facilities, battalion aid stations, or in-theater facilities.

Chemistry data are useful for identifying laboratory confirmed cases of illness. However, cases where a physician chooses to treat presumptively without laboratory confirmation are not captured. Clinical practice with regards to chemistry testing varies between providers and facilities. Examples of situations where chemistry testing may not be performed include confirmatory tests for patients with influenza-like illness symptoms, or patients with superficial infections who are treated presumptively.

Background

The EpiData Center Department (EDC) of the Navy and Marine Corps Public Health Center (NMCPHC) was funded by the Department of Defense (DOD) Global Emerging Infections Surveillance and Response System (GEIS) to evaluate the Health Level 7 (HL7) chemistry data source for its usefulness in public health surveillance. While HL7 data also includes radiology, anatomic pathology reports and pharmacy transactions, laboratory results, including both chemistry and microbiology, were identified as the most useful type of data for improving military public health surveillance activities. Therefore, extensive work has been done to examine the laboratory datasets, determine their completeness and reliability, identify areas for improvement as needed for surveillance, and establish methods for the surveillance of specific conditions. Available data fields have changed throughout the course of HL7 development; this document reflects variables available as of July 2012. This technical document provides a history of the HL7 chemistry database and its contents, explains the creation of chemistry records, describes the pathway of data from the healthcare provider to the EDC, provides a detailed descriptions of all variables within the database, and assesses the database's strengths and limitations.

Records for all DOD military service members (Army, Navy, Marine Corps, Air Force, Coast Guard, and US Public Health Service), overseas civilian personnel, Tri-Care eligible dependents, and others who undergo chemistry testing at a military treatment facility (MTF) are included in the HL7 chemistry dataset. Chemistry records for the Department of Navy (DON) have been collected at NMCPHC since 01 May 2004. Due to additional GEIS funding to support DOD Pandemic Influenza surveillance, HL7 chemistry laboratory data for the entire DOD have been collected at NMCPHC since 21 July 2006.

Public Health Surveillance Applications

The HL7 chemistry database has been used to support DON and DOD Preventive Medicine activities since 2005. Support includes identifying cases of reportable diseases, analysis of case burden, and estimation of occupational exposures as indicated by organic biomarkers such as lead and cadmium. Epidemiologic analysis of HL7 chemistry data focus on identifying trends of illness or condition. Positive laboratory results are matched to encounter records to identify how specific laboratory tests relate to clinical encounters. Alternatively, International Classification of Diseases, 9th Edition (ICD-9) codes in encounter data are also matched with HL7 chemistry laboratory data to identify how specific encounters relate to laboratory testing. Many variables in a data-based disease surveillance model depend on clinical practice. These approaches will provide a more comprehensive description of how laboratory tests and results compare with information provided in encounter data.

Additional public health applications involve investigation into the types and subtypes of disease and their impact on public health measures or policies. For example, knowing the subtype of a disease can assist in the predictions for the upcoming season, or aid in the development of

vaccinations. Applications of these data are not limited to traditional medical event surveillance. These data can fill a significant gap in the DOD's ability to track conditions identified through chemistry laboratory testing.

Data Origination and Flow Process

Figure 1 provides an overview of the flow of information in the HL7 chemistry database. Generation of HL7 laboratory data originates at the point of care, when a patient is seen by a provider in a fixed-MTF and the provider orders a laboratory test via the Composite Health Care System (CHCS). Generally, the provider selects from a list of tests. The list in CHCS is generated by the MTF laboratory Management Information Department (MID) based on laboratory tests that can be run at the local facility or outside laboratories that have contracts with the facility. Laboratory tests available to order vary by facility. Additionally, the same test may be performed at different facilities with different names based on how they are set up in CHCS. Laboratory orders can also be entered into CHCS from providers outside the military healthcare system or through a referral from a non-CHCS participating military provider (e.g. ship-board clinic, battalion aid station) if the patient provides a laboratory order slip. Depending on the patient's status, specimens can be drawn within the hospital ward and sent to the laboratory (inpatient), or can be drawn directly at the laboratory during clinical encounters (outpatient). A laboratory test can be performed on site, or may be outsourced if the local facility is unable to perform the test indicated. When a test is completed, results are entered into CHCS either automatically by the machine performing the test or manually by the laboratory technician. Results are then certified by a laboratory technician or supervisor. After certification in CHCS, a script is run to generate an HL7 message for that test result. The HL7 message is then archived and batched with other HL7 messages on the local CHCS host. At least once a day, these HL7 messages are forwarded to Defense Health Services System (DHSS) main servers. Once forwarded and receipt is verified by DHSS, HL7 messages at the local host are deleted. These records are then retrieved from the main servers and parsed into a database design four times a day. The EDC receives flat file extracts of the raw parsed data from DHSS on a daily basis using a secure connection, as dictated by the Interface Control Document (ICD).

Data Structure and Analysis

Structure

HL7 chemistry data are retrieved by the EDC in a standard, pipe-delimited flat file from DHSS. Variables provided are described in detail in a subsequent section of this document (field observations). Multiple tests can be ordered together, called panels, and these rows can be grouped together using ACCESSION NUMBER, SPECIMEN SOURCE, and SET ID.

When a test is canceled, inaccurate, or misread, the details are recorded within the TEST RESULT or the RESULT NOTES field; a new row is not created. When a test is reordered, a separate row is generated to show the newest results, and references the previous test.

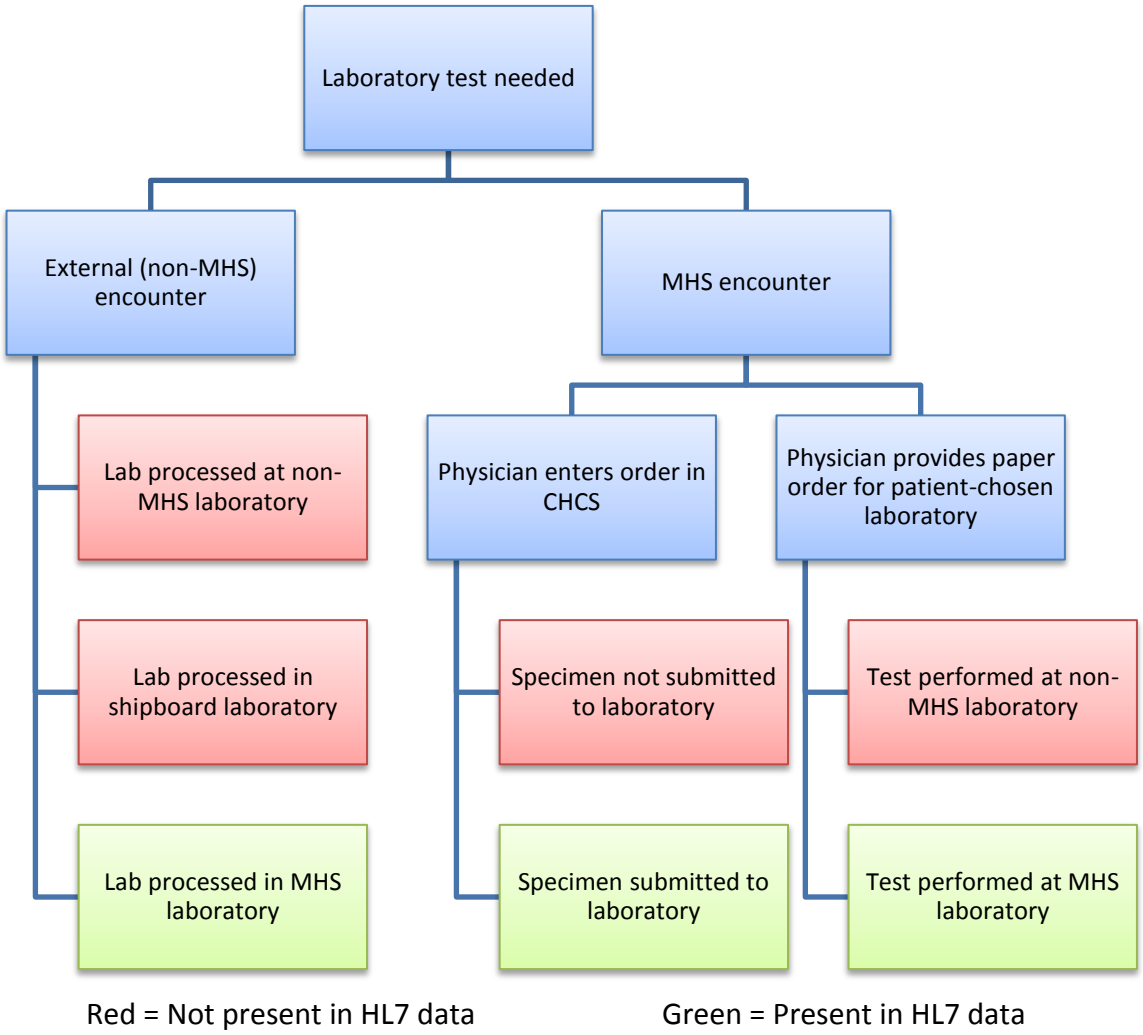
Analysis

Most military laboratory facilities can perform routine chemistry testing, such as Complete Blood Counts (CBCs), Chem 7s, Rapid Plasma Reagents, and Hepatitis panels. More specialized tests are usually outsourced to a private laboratory (e.g. Western Blots). If a test is outsourced, MTFs are required to ensure those results are entered into CHCS for clinical evaluation by the provider following the receipt of results from the contracted laboratory. However, all MTFs may not be compliant with this policy.

Laboratory tests that are not traditionally classified as chemistry tests may be found in the chemistry dataset due to MTF-specific situations. This may occur if an MTF acquires microbiology specimens through the chemistry section of the laboratory. This practice is more common in smaller laboratories without a microbiology-specific sector, which may perform non-chemistry tests in the chemistry laboratory. Therefore, when searching for a particular outcome, both laboratory datasets (chemistry and microbiology) should be included to ensure capture of all applicable test results.

A laboratory test can be used as a diagnostic tool but is not required for many illnesses, and may not estimate the true burden of disease. For instance, in some cases a provider may identify and treat patients based on clinical symptoms alone. Other times, based on epidemiologic characteristics such as extent of disease in the community, seasonality, or disease characteristics, the disease may be identified without laboratory confirmation. Thus, a disease surveillance model based on laboratory results alone may skew illness patterns, unless these and other factors are taken into consideration. Figure 1 illustrates the data flow process.

Figure 1. HL7 laboratory data flow process



Several things should be taken into consideration before analysis of a particular outcome or disease. It is critical to understand the pathogen or disease, its clinical presentation, provider practices with regard to treatment and laboratory testing for the disease/pathogen, available testing methods, and laboratory testing or outsourcing procedures. It is also important to note that test names and results are not standardized in CHCS or in the HL7 data, but are often standardized at each MTF. The CHCS administrator at an MTF may set up drop-down menus and panels based on the facility's capabilities. Multiple test names and result formats may be associated with a particular disease depending on the testing method and on recording practices of the facility. Therefore, in conducting a query of particular tests in the HL7 chemistry table, misspellings and abbreviations should be included. Formatting variations should also be considered, including spacing and use of periods. Multiple fields may be searched for flagged tests associated with a particular outcome as well. The subset of data created for the disease of interest should then be reviewed to assess completeness and reliability of the data with regard to

that particular outcome.

To re-evaluate and assess the HL7 chemistry dataset and provide insight into data fields as described below, data were extracted for DON beneficiaries from the EDC's HL7 chemistry database from calendar year (CY) 2011. All data were reviewed and analyzed to ensure the dataset addresses the disease surveillance needs of the Navy and Marine Corps. The data extract reflects information from the personnel and their dependents of US Marine Corps, US Navy, or any beneficiary whose test results were certified in the CHCS at a US Navy MTF. The same procedure was conducted with the DOD HL7 data to identify characteristic differences between the DOD and DON data.

Key Fields for Public Health Surveillance

Data were pulled for all DOD medical beneficiaries on July 18, 2012 from the HL7 chemistry database. All data were reviewed and analyzed in order to modify the datasets to more accurately address the disease surveillance needs of the EDC. Methods for identifying duplicate and unique records were established.

Duplicates

Within the HL7 chemistry dataset, there are several ways in which duplicate records can be identified. Duplicate rules described here should be checked against project objectives to ensure applicability. True duplicates are records in which all fields have identical values. True duplicate records meeting this criterion should be eliminated. Each record that remains after removing true duplicates is considered a unique record; there is at least one variable value different than all other records in the database.

Unique Patient/Specimen

Unique patients are identified in the HL7 chemistry data through a combination of SPONSOR ID and FMP. This combination creates a unique identifier that can be used to track individual patients through all chemistry records. There is a variable called PATIENT ID; however, the field is not complete, consistent, or reliable as a source of identifying patients within or across databases. It is important to note that it is possible for individuals to have two separate SPONSOR IDs over time. For example, if the child of a sponsor becomes active duty, then that child will have his/her own SPONSOR ID. Each unique patient can have multiple laboratory orders in the HL7 chemistry data.

A unique specimen is defined as each sample collected from each SPECIMEN SOURCE from each person. A unique ACCESSION NUMBER is assigned to each unique specimen collected. Patients may have multiple samples taken at the same time. Each sample would have a different ACCESSION NUMBER, even if many of the other fields are the same. ACCESSION NUMBER may be reused by the MTF, but when combined with SPONSOR ID, FMP, and ORDER EFFECTIVE DATE, the combination is unlikely to be reused. Since multiple tests can be performed on the same sample, there can be several records with the same ACCESSION NUMBER.

Test Definition

The TEST ORDERED and the TEST NAME fields are used to identify the test performed. Variation between computer coding within the regional CHCS and the capabilities of each laboratory causes deviations and non-standardization of test names. It is important to consider this in analysis of these data. For example, with influenza, a laboratory test could be defined by taxa (order, family, subfamily, genus, and species) or test type (antibody staining test, convalescent test). Each influenza test has a different specimen type (i.e. serum, nasopharyngeal wash), result type (i.e. numeric versus positive/negative), and timeframe which determines each

testing method (Table 1: Influenza Diagnosis Table). All of this information should be taken into consideration when flagging a test in the HL7 data. Table 2 provides an example of an HL7 data feed for one patient, showing ACCESSION NUMBER, TEST NAME and TEST RESULT.

Table 1. Influenza Laboratory Diagnosis Table. CDC 2007.

Influenza Diagnostic Table				
Procedure	**Influenza Types Detected**	**Acceptable Specimens**	**Time for Results**	**Rapid result available**
Viral culture	A and B	NP swab[2], throat swab, nasal wash, bronchial wash, nasal aspirate, sputum	3-10 days[3]	No
Immunofluorescence DFA Antibody Staining	A and B	NP swab[2], nasal wash, bronchial wash, nasal aspirate, sputum	2-4 hours	No
RT-PCR[5]	A and B	NP swab[2], throat swab, nasal wash, bronchial wash, nasal aspirate, sputum	2-4 hours	No
Serology	A and B	paired acute and convalescent serum samples[6]	>2 weeks	No
Enzyme Immuno Assay (EIA)	A and B	NP swab[2], throat swab, nasal wash, bronchial wash	2 hours	No
Rapid Diagnostic Tests				
Directigen Flu A[7] (Becton-Dickinson)	A	NP wash and aspirate	<30 minutes	Yes
Directigen Flu A+B[7,9] (Becton-Dickinson)	A and B	NP swab[2],aspirate, wash; lower nasal swab; throat swab; bronchioalveolar lavage	<30 minutes	Yes
Directigen EZ Flu A+B[7,9] (Becton-Dickinson)	A and B	NP swab[2], aspirate, wash; lower nasal swab; throat swab; bronchioalveolar lavage	<30 minutes	Yes

FLU OIA[4,7] (Biostar)	A and B	NP swab[2], throat swab, nasal aspirate, sputum	<30 minutes	Yes
FLU OIA A/B [7, 9] (Biostar)	A and B	NP swab[2], throat swab, nasal aspirate, sputum	<30 minutes	Yes
XPECT Flu A&B[7,9] (Remel)	A and B	Nasal wash, NP swab[2], throat swab	<30 minutes	Yes
NOW Influenza A[8,9] (Binax)	A	Nasal wash/aspirate, NP swab[2]	<30 minutes	Yes
NOW Influenza B[8,9] (Binax)	B	Nasal wash/aspirate, NP swab[2]	<30 minutes	Yes
NOW Influenza A&B[8,9] (Binax)	A and B	Nasal wash/aspirate, NP swab[2]	<30 minutes	Yes
OSOM® Influenza A&B[9] (Genzyme)	A and B	Nasal swab	< 30 minutes	Yes
QuickVue Influenza Test[4,8] (Quidel)	A and B	NP swab[2], nasal wash, nasal aspirate	<30 minutes	Yes
QuickVue Influenza A+B Test[8,9] (Quidel)	A and B	NP swab[2], nasal wash, nasal aspirate	<30 minutes	Yes
SAS Influenza A Test[7,8,9]	A	NP wash[2], NP aspirate[2]	<30 minutes	Yes
SAS Influenza B Test[7,8,9]	B	NP wash[2], NP aspirate[2]	<30 minutes	Yes
ZstatFlu[4,8] (ZymeTx)	A and B	throat swab	<30 minutes	Yes

1. List may not include all test kits approved by the U.S. Food and Drug Administration
2. NP = nasopharyngeal
3. Shell vial culture, if available, may reduce time for results to 2 days
4. Does not distinguish between influenza A and B virus infections
5. RT-PCR = reverse transcriptase polymerase chain reaction
6. A fourfold or greater rise in antibody titer from the acute- (collected within the 1st week of illness) to the convalescent-phase (collected 2-4 weeks after the acute sample) sample is indicative of recent infection.
7. Moderately complex test – requires specific laboratory certification.
8. CLIA-waived test. Can be used in any office setting. Requires a certificate of waiver or higher laboratory certification
9. Distinguishes between influenza A and B virus infections

Disclaimer: Use of trade names or commercial sources is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or the Department of Health and Human Services.

Table 2. Example of an HL7 data-feed for one patient showing ACCESSION NUMBER, TEST NAME, AND TEST RESULT

ACCESSION NUMBER	SET ID	TEST ORDERED	TEST NAME	TEST RESULT
070101 – ACH – 045	1	CBC W/O DIFF	HEMATOCRIT	36.9
070101 – ACH – 045	2	CBC W/O DIFF	HEMOGLOBIN	12.7
070101 – ACH – 045	3	CBC W/O DIFF	MCH	28.7
070101 – ACH – 045	4	CBC W/O DIFF	MCHC	34.5
070101 – ACH – 045	5	CBC W/O DIFF	MCV	83.2
070101 – ACH – 045	6	CBC W/O DIFF	MPV	7.9
070101 – ACH – 045	7	CBC W/O DIFF	RBC	4.43
070101 – ACH – 045	8	CBC W/O DIFF	WBC	5.5
070101 – ACH – 045	9	CBC W/O DIFF	PLATELET CT	255
070101 – ACH – 047	1	CHEM 7	CREATINE	0.6
070101 – ACH – 047	2	CHEM 7	UREA NTROGEN	13
070101 – ACH – 047	3	CHEM 7	GLUCOSE FASTING	172
070101 – ACH – 047	4	CHEM 7	SODIUM	136
070101 – ACH – 047	5	CHEM 7	POTASIUM	3.6
070101 – ACH – 047	6	CHEM 7	CHLORIDE	103
070101 – ACH – 047	7	CHEM 7	CARBON DIOXIDE	23

Laboratory Test Result

Due to the structure of the laboratory datasets, the result or its subgenus (or subtype) could be recorded in multiple fields. Most test results for chemistry are seen in the TEST RESULT field. When a result is not present within this field, usually identified by terms such as 'See Results', 'See Note', etc., a search within the RESULT NOTES could indicate a positive or negative result. For those tests with a "Positive" result, it is recommended to view the TEST NAME and the SET_ID = 1 to identify the positive agent.

An entry could have multiple results per ACCESSION NUMBER, either through errors, or multiple tests performed. Therefore, it is important to review the records based on the most recent results certified by the laboratory (CERTIFY DATE), the most certain RESULT STATUS OBX (Corrected, Final, then Pending), and the SET ID to select the most accurate results.

Date/Time References

Figure 2 illustrates the five date fields with associated time fields in the chemistry laboratory dataset. The MSG DATE and MSG TIME fields show the most accurate timeframe of when the data were available at CHCS. Data pulls for epidemiologic projects primarily use the ORDER EFFECTIVE DATE to capture the timeframe when the patient is likely to be symptomatically ill, which often leads to tests ordered to assist with diagnosis. ORDER EFFECTIVE DATE is the date the provider requests the laboratory test be performed. It is possible for a provider to enter a laboratory test order to be performed in the future; this may be common practice for chronic illnesses or long term inpatient visits. It is also possible that the patient did not immediately report to the facility to have the specimen collected. Therefore, the appropriateness of using ORDER EFFECTIVE DATE should be considered. An alternative proxy may be the sample COLLECTION DATE, which is the date that the specimen was taken from the patient.

Figure 2. Timeline of dates in HL7 chemistry

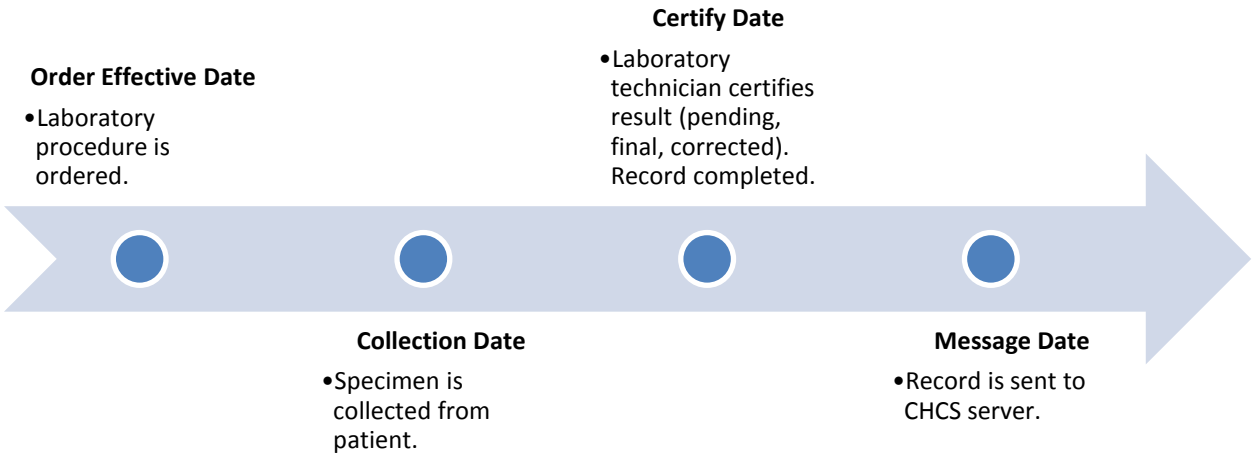


Table 3 reflects methods used to determine timeframes within the HL7 chemistry database. Each date and time pair references one action towards the completion of the laboratory test. A time span was created when comparing one date/time variable to another date/time variable.

Table 3. Methods used to determine timeframes in HL7 data

Date/Time Variables	Interpretation
Message – Order Effective	Determines timeframe from when test was ordered to when an HL7 message was generated at the local CHCS host
Order Effective – Date of Birth	Age of member at date of test order
Collection – Date of Birth	Age of member at date when specimen was extracted.
Certify – Order Effective	Timeframe from test ordered to results
Certify – Collection	Timeframe from when sample was taken to results

Strengths

Timeliness

DHSS includes several date and time fields in the data file provided to the EDC, including: MSG DATE/TIME, ORDER EFFECTIVE DATE/TIME, REQUESTED DATE/TIME, COLLECTION DATE/TIME, and CERTIFY DATE/TIME. A timeline of dates is provided in Appendix E. To assess the timeliness of the data, the ORDER EFFECTIVE DATE (date the order was placed into CHCS by the provider) was compared to the MESSAGE DATE (date the HL7 message was generated by CHCS) to estimate the time between patient encounter and receipt of data at DHSS.

On average, it takes about two days for an HL7 message to be generated and received by the EDC. Future analysis and assessment should better define lag times in relation to particular MTFs, testing, or outcomes of interest.

Completeness

Records are received from the majority of shore-based fixed MTFs connected to the CHCS, but gaps in data may exist. Gaps in data may occur due to server failure at location or due to functional errors. The completeness of individual fields varies and the characteristics of each are described in detail in the field observations section of this document. In general, some fields of particular interest, such as SPONSOR ID, FMP, and SERVICE are highly populated due to the business rules of CHCS. Entries are only received by DHSS once the test has been completed and certified, which increases the accuracy of data for outcome analysis and surveillance.

Organization and Structure

Data are organized in a useful manner. Data fields of highest importance for surveillance are complete and contain expected variables. Test results are located on single rows of the data table, which limit the need for restructuring. Single line entries allow for flexibility with stratifications, filtering, and aggregating. Timelines can then be determined within each test from the date and time variables.

Limitations

Completeness

The HL7 infrastructure at DHSS was built using pilot funds, but as of March 2009 is a functional program. Initially, a temporary network was created to capture HL7 messages when they were sent from the CHCS host to DHSS. Until the program became formal, no back-up system existed. When the DHSS node fails, HL7 messages may be lost and those that had been sent may not have been retrievable unless the network outage was planned for in advance. Gaps may exist in the data received at NMCPHC, though are limited. Several of the identified data fields of public health interest are highly populated, but others are not. The completeness of each data field, as described later in this document, should be considered.

Incomplete demographic information can limit the ability to describe the characteristics of an outcome. Complementary databases, such as personnel data, may need to be utilized to supplement HL7 data for more complete and accurate demographic information.

Inclusion

The chemistry data only includes records generated by MTFs that use CHCS servers. MTFs that do not use CHCS (including forward deployed clinics, ship-board clinics, and battalion aid stations) and tests performed by outside laboratories (outsourced or external provider/laboratory) are not captured in this dataset. Records for tests ordered by these clinicians are present in the dataset if a service member is referred to an MTF that uses CHCS. Additionally, the severity of illness or lack of on-location capabilities may require the service member to be transferred from a non-CHCS MTF to a shore-based CHCS facility. Comparatively, should a beneficiary seek care outside the MHS, such as the use of Tri-Care Prime, his/her records would not be captured within CHCS or HL7 unless the laboratory testing was completed at an MTF laboratory.

Generalizability

Data are generated from the medical laboratory records of a highly specific patient population – military service members and their dependents – which differs from the general US population in many ways, including average age, gender distribution, physical fitness, and health status. Further, this population has universal access to medical care, which is not true of the US population. These differences limit the comparability to the general US population. Extra precautions need to be taken when extrapolating data to larger populations and also when comparing the disease rates and trends of the military and non-military populations.

Inconsistency

The format of CHCS allows for free text fields (e.g., TEST NAME, TEST RESULT, RESULT NOTES), resulting in variations between entries, and difficulties in HL7 data interpretation. Free text fields limit the ability to determine specific test type (e.g. enzyme immunoassays or culture), test results, and reference ranges. Non-standardized naming and result conventions hinder the ability to use standardized syntax. However, methods have been developed at the EDC to overcome these barriers.

Inconsistencies between MTFs limit data use. Tests performed are based on the capability of each laboratory. The MTF laboratories determine testing capabilities and control what providers may order on their CHCS laboratory order screens. Also, TEST NAME and TEST RESULT options in CHCS (pull down menus) are determined and controlled at the local CHCS host, allowing variation. The variance increases the need for extensive coding to determine case definition, ensuring comprehensive case capture.

Data Interpretations

While these data may be useful to identify laboratory confirmed cases of illness, not all cases will be present if physicians elect to treat patients presumptively without laboratory confirmation. Clinical practices, and reliance on symptomatic examination versus diagnostic testing, may vary between providers. For example, due to the volume of cases during the influenza season, providers may be less likely to test each individual who presents with influenza-like symptoms.

All Data Fields (Variables)

Automatically Populated Fields

There are several types of automatically populated fields in the chemistry data that are created at different levels, including: the CHCS host, each MTF (a division of the host), and each laboratory work section.

When a facility registers within CHCS, several variables are created for use in identification of the facility per DHSS. These variables include: PERFORMING DMIS ID, PERFORMING DMIS FACILITY NAME, PERFORMING FACILITY SERVICE, REQUESTING DMIS ID, REQUESTING FACILITY NAME, and REQUESTING FACILITY SERVICE. Additionally, DHSS populates the SET ID and SERVICE variables.

Each patient or dependent is registered in the Defense Eligibility Enrollment Reporting System (DEERS), which feeds into CHCS. When a patient presents at a medical facility, the SPONSOR ID (usually the social security number) is entered and their name is chosen from a drop down list. The following patient demographic fields are automatically populated after this selection, if they were entered when the patient was registered in DEERS: DATE OF BIRTH, ETHNICITY, FMP, GENDER, MARITAL STATUS, PATIENT CATEGORY, PATIENT ID, RACE, and SPONSOR ID. If these data are not present in the system, a designated unknown value is entered, and, therefore, there are no missing values in these fields. Administrative personnel at the MTF have the ability to edit records at the time of visit. It is the sponsor's responsibility to update his/her family's DEERS information when they report to a new duty station or when demographic information changes.

As records are created, edited, and completed, several variables are created by the CHCS clock when a specific action is taken, such as order entry or certification of a test result. These variables include: ORDER EFFECTIVE DATE/TIME, SPECIMEN COLLECTION DATE/TIME, REQUESTED DATE/TIME, and CERTIFY DATE/TIME. These can be changed if necessary by the laboratory staff, but is not common practice.

MSG DATE, MSG ID, MSG TIME, and MSG SENDING FACILITY are created and assigned when the message (record) is sent to the CHCS server. The MSG ID value is created when the certified laboratory record is processed by the CHCS server at the local level for batching. The time and date are also automatic values created by the system. The MSG SENDING FACILITY is a code which identifies the server and facility creating the message.

The CHCS host or MTF automate variables prior to a physician selecting a test from the drop-down list or lookup tables. Should a physician choose a test from the TEST ORDERED field, a variation of tests within the TEST NAME could be selected. Based on the physician's request, the SPECIMEN SOURCE, BODYSITE COLLECTION SAMPLE, CPT CODE DATA, and NO

OF CPT CODES are selected for each TEST NAME. These can be changed if necessary by the laboratory staff.

The laboratory results and each reference range can be entered by the laboratory staff, automated by the Laboratory CHCS Coordinator, or by the specific laboratory equipment used. Specific models of laboratory equipment can be specialized to indicate the normal ranges of values, label abnormalities, and create syntax per each specimen's result. Depending on each laboratory facility and equipment used, the TEST RESULT, RESULT NOTES, REFERENCE RANGE, ABNORMAL FLAG, SENSITIVE RESULT FLAG, and UNITS OF MEASURE can be automated or manually entered by staff.

Variable Descriptions

Observations are based on DOD data. Within the chemistry dataset, there are 55 variables. Frequency distributions for CY2011 were run on select data fields from the chemistry database to describe completeness. These fields are presented in alphabetical order.

ABNORMAL FLAG

ABNORMAL FLAG is a string field which indicates the status of the result. The majority of the entries have missing values. There are three different values which can be listed: '/', H (High), and L (Low).

ACCESSION NUMBER

The format of the ACCESSION NUMBER is a combination of the 1) date in an YYYYMMDD format, 2) a two or three character alpha setting, and 3) a numeric listing of how many tests of that specific type were run in one day. The last numeric digits can range from 1 to 9999. ACCESSION NUMBERS are created for each unique biological sample collected from the patient. Different chemistry tests from the biological sample can have the same ACCESSION NUMBER. These numbers could be recycled throughout a day's time, and should not be used solely to identify a record. They may be used to determine tests ordered per patient in conjunction with the SPONSOR ID, FMP, and ORDER EFFECTIVE DATE.

BODYSITE COLLECTION SAMPLE

The BODYSITE COLLECTION SAMPLE refers to the place on the body where the specimen is collected from the patient. This field is used with SPECIMEN SOURCE to determine where the sample is taken. A patient can have numerous samples taken from one area (i.e. a lung can have numerous biopsy specimens, thus having a different ACCESSION NUMBER for each specimen). But, like SPECIMEN SOURCE, it can be used to determine if proper protocol was used for a test, or can be used to determine the type of test performed (i.e. CBC would not have urine or bladder as a sample type).

CERTIFY DATE

The CERTIFY DATE is the date when a laboratory technician certifies the results into CHCS, or makes changes within the system. Unlike the ORDER EFFECTIVE DATE, there can be deviations between the values within SET ID, due to different test run dates. The CERTIFY

DATE is formatted YYYYMMDD and contains limited missing values. All entries have valid dates. The values of the timeframe are between ORDER EFFECTIVE DATE and MSG DATE.

CERTIFY TIME

This field represents the time component of the CERTIFY DATE and is formatted using a standard 24 hour clock. The possible values are from 0000 to 2359, and all times are valid entries.

CLINICAL COMMENTS

The CLINICAL COMMENTS is a free text field which allows the provider or laboratory technician to add additional information regarding the patient's symptoms, CAP[1] survey information, contact phone numbers, specimen media, or instructions on test procedures. This field is not primarily used in case definition, but is added to eliminate superfluous entries.

COLLECTION DATE

The COLLECTION DATE is the date when the specimen is extracted from the patient. The value for this entry should be between the values of the ORDER EFFECTIVE DATE and the CERTIFY DATE. The COLLECTION DATE is formatted YYYYMMDD and includes limited missing values. All records without missing values have valid dates.

Since the field approximates the time that the laboratory sample is collected, it may be useful for analysis. It can be used for time analysis between the specimen collection and test result certification. By knowing the timeframe of each test conducted, an approximation of the type of test used can be determined.

COLLECTION TIME

As with COLLECTION DATE, the COLLECTION TIME is the time when the specimen is extracted from the patient, and follows a standard 24-hour clock. Unlike ORDER EFFECTIVE TIME, the timeframe is from 0001 to 2400. All times are valid entries. There are no missing values.

CPT CODE DATA

The CPT CODE DATA is an alphanumeric field which identifies a particular test by the Current Procedural Terminology (CPT) code. The CPT code is identified per the American Medical Association, and describes medical, surgical, and diagnostic procedures. This is designed to communicate uniform information about medical services and procedures between physicians, coders, patients, accreditation organizations, and payers for administrative, financial, and analytical purposes.

The variable format is #####\##\AD. The first group of characters defines the CPT code used within the chemistry dataset. The second portion is a modifier code which indicates the accession area and work element. The values observed in the chemistry data are defined as: 26 –

[1] College of American Pathology (CAP) conducts surveys for quality assurance within the laboratories.

Professional/Pathologist, 32 – Mandated Service (MTF performs laboratory for a branch clinic), 90 – Reference Laboratory Service (e.g. LabCorp), 91 – Repeat Clinical Diagnostic Procedure (multiple tests for subsequent results). The value of 00 is present but is undefined by the reference. AD stands for Active Duty.

A regional CHCS site maps a CPT code to a particular methodology or technique. CPT codes are assigned at various levels to the CHCS test files when the laboratory sets up the procedure. All tests that do not have a specific CPT code may be given unlisted procedure/service codes defined for the specific types of test (immunology, chemistry, microbiology, hematology, etc.).

DATE OF BIRTH

The DATE OF BIRTH field (DOB) is included in the format YYYYMMDD. It is possible to have inaccurate values for DOB. If the DOB is unknown and the year is, then CHCS enters zeros for the month and day. Not all dates for this field are valid (i.e. dates with a year in the early 1900s or a date with a year in the future). This field is required within CHCS; therefore, there are no missing values.

DHSS LOAD DATE

DHSS LOAD DATE indicates the date when DHSS loads the data from the central CHCS server. The field is used to determine the timeliness of reporting and to identify lags in reporting times from certain MTFs. The format is YYYYMMDD and there are no records with missing values.

DHSS LOAD TIME

Time component of the DHSS LOAD DATE field, and is formatted: HHMM. The values present in the data are 0300, 1000, 1600, and 2000, and there are no records with missing values.

ETHNICITY

ETHNICITY is an alpha numeric field with six possible values; 1=Hispanic, 2=South Eastern Asian, 3= Filipino, 4=Other Asian Pacific Islander, 9=Other, and Z=Unknown. The highest group is Other with 49.0%. About 44.0% of records from chemistry are categorized as Unknown. There are no missing values. Of these categories, the Hispanic ethnicity makes up 3.7% of the records within the DOD extract. These results seem to indicate that the field of ETHNICITY may be self-identified and not consistently reported. Those entries which are not reported are labeled as Unknown. The Unknown responses are assumed to be pre-populated in order to eliminate blanks within the database. It limits the ability to identify disease trends in minority groups and to identify diseases that have a disproportionate burden on these groups.

FMP

FMP is the family member prefix that designates the relationship of the patient to the sponsor. The distribution of FMP among the records is as expected, with the highest numbers recorded as 1-3, 20, and 30 which are values that correspond to first, second, and third child of sponsor (FMP=1-3), the sponsor (FMP=20), and spouse of sponsor (FMP=30). All records have a value for FMP.

GENDER

There are three values possible for the GENDER field; M=Male, F=Female, X=Unknown. Of the records included from chemistry, there are no missing values, and all are either coded as M or F.

MARITAL STATUS

There are nine values for MARITAL STATUS: A=Annulled, D=Divorced, I=Interlocutory Decree, L=Legally Separated, M=Married, S=Single/Not Married, W=Widow or Widower, Z=Unknown. There are no missing values for records in the chemistry dataset. Nearly 42% of all records are categorized as Married, followed by Unknown (40.4%) and Single/Not Married (12.7%).

MEPRS CODE

The MEPRS (Medical Expense and Performance Reporting System) CODE is a four letter code that indicates where within the MTF the laboratory order was entered. The first letter indicates the most general area and translates as: A=inpatient, B=outpatient, C=Dental, D=ancillary, E=support services, F=special programs, and G=medical readiness. It is advised to obtain an up-to-date list of all possible codes. The chemistry dataset does not have missing values because it is automatically populated when the record is created. This field is useful for tracking where people are seen within the MTF. It can indicate ambulatory care, special dialysis clinics, the maternity ward, etc. which can affect the interpretation of the data. The majority of records present in the chemistry dataset have a MEPRS code that begins with B.

MSG DATE

This field is formatted YYYYMMDD. There are no missing values and all are valid dates. This date approximates the transaction time between the MTF and the regional CHCS site, but it can vary based on location. Some MTFs send messages in batches, therefore the time or date portions may not correlate to the actual transaction time.

MSG ID

The Message ID (MSG ID) is an alpha numeric code assigned to each batch of messages based on when the message is sent from CHCS to the server. The MSG ID is not unique to each record; each batch of messages is assigned one MSG ID. The MSG ID format varies by MTF and may include numbers, letters or numeric code that identifies the MTF, or it can identify the function of the message (i.e. RESCHED-057342).

MSG SENDING FACILITY

This field is formatted as AA###. This field allows analysts to identify and track problems that arise in the transfer of messages from the MTFs to DHSS and the EDC. There are no missing records within this dataset.

MSG TIME

The MSG TIME is the time when the message is sent from the MTF to the regional CHCS site, and follows a standard 24-hour clock. The numbers range from 0001 to 2359. There are no recorded times for 0 or 2400. All times are valid entries. There are no missing values.

NO OF CPT CODES

The NO OF CPT CODES is a numeric field which lists the number of CPT codes used for each test performed. Over three-quarters of the entries only have one CPT code associated with them. The number of CPT codes is determined at each regional location.

ORDER EFFECTIVE DATE

The ORDER EFFECTIVE DATE is the date that the laboratory order enters CHCS. It is different from the MSG DATE since the MSG DATE is generated after the laboratory results are certified. The ORDER EFFECTIVE DATE more accurately approximates when the laboratory test is actually ordered. The ORDER EFFECTIVE DATE is formatted YYYYMMDD and includes two missing values. Since the field approximates the time that the laboratory test is ordered, it may be used to identify when the patient presented with clinical symptoms necessitating the test, and to allow for time analysis between the order dates and sample collection date. Additionally, it may assist in determining the duration until the completion of the test, to determine which type of test is used, and to identify time lags between when the test is ordered and when data are available for analysis at the EDC.

ORDER EFFECTIVE TIME

The field represents the time component of the ORDER EFFECTIVE DATE and is formatted using a standard 24-hour clock. Unlike MSG TIME, this timeframe includes values for 0000. The range present is 0000 to 2359, and all times are valid entries.

ORDER NOTES COMMENTS

The ORDER NOTES COMMENTS is a text field which allows the provider to include notes or comments that accompany the test ordered. Multiple Patient Notes (NTEs) can be in the same field and are separated by a "~" symbol. Repetitive NTEs are separated by a blank space. This field is not currently populated in the dataset.

ORDER NUMBER

The ORDER NUMBER is a numerical code with eleven digits (xxxxxx-xxxxx) unique to each order but not unique for each record. These numbers are unique for each location, and are not circulated. The first set of numbers is the date, and the last 5 numbers are consecutive for tests provided at that specific location. An order can have multiple records that correspond to changes made to the order (i.e. changes in test, cancellations), or refer to multiple parts of the test (e.g. results for influenza A and influenza B). All changes appear as individual records with the same ORDER NUMBER. It is a plausible way to track a patient but it is not useful for identifying unique records.

ORDERING PROVIDER

The ORDERING PROVIDER field indicates the name of the ordering physician. It has three components each separated by a ",": Last Name, First Name, Middle Initial. It is structured to facilitate analysis but could be separated if necessary.

PATIENT CATEGORY CODE

The patient category code (PATCAT CODE) is an alphanumeric code that indicates the patient's status with the uniformed services. The first letter of the code refers to the branch of service of the Sponsor (A=Army, B=National Oceanic and Atmospheric Administration, C=Coast Guard, F=Air Force, K=other beneficiaries of the federal government, M=Marine Corps, N=Navy, P=US Public Health Service, R=NATO recipient). It is followed by two digits corresponding to the patient relationship to the sponsor. For example: A11=Army Active Duty Member, A41=Army Dependants of Active Duty, etc. The most frequently reported PATCAT CODE in all branches is 41 (dependent of active duty), 11 (active duty), and 43 (dependent of retired sponsor), respectively. A complete list should be obtained from DOD resources. Less than 1% of records are missing PATCAT CODES in the chemistry database.

PATIENT ID

The PATIENT ID is intended to serve as a unique identifier for each patient. The format for PATIENT ID is a nine digit numeric listing. In the Interface Control Document (ICD) provided by DHSS, it states that the PATIENT ID is the patient's SSN when available. Few records have a null PATIENT ID field. PATIENT ID cannot be validated based on observations of the data received by the EDC. The SPONSOR ID in conjunction with FMP can be used as a substitute unique patient identifier. Importing this field in character format can prevent the loss of leading zeros.

PERFORMING DMIS FACILITY NAME

This field is the text translation of the DMIS ID provided in the PERFORMING DMIS ID field. This field indicates the laboratory facility name where the specimen was sent, not where the specimen was collected from the patient. This field is assigned by DHSS at the request of the EDC. The translation of the DMIS code on the official list is often more accurate than the Performing Facility field in CHCS. Use of this field allows for more accurate analysis of geographic information. Since the field is also a translation of the performing facility field in CHCS, it will be missing when that variable has a missing value.

PERFORMING DMIS ID

The PERFORMING DMIS ID is a four digit code assigned by the DOD to all units at all installations in order to uniquely identify them. This field indicates the DMISID where the specimen was sent, not where the specimen was collected from the patient. The EDC provided an official DMIS list to DHSS for the purpose of creating this variable. DHSS translated the Performing Facility field within CHCS to its assigned DMIS code. This code allows for grouping of MTFs based on geographic location, as well as to identify parent/child relationships between installations. Since this field is calculated based on the performing facility field, all records missing a value for that field will be missing a value for the PERFORMING DMIS ID

field. Importing this field in character format can prevent the loss of leading zeros, which may produce complications when producing summary statistics.

PERFORMING FACILITY SERVICE

The PERFORMING FACILITY SERVICE field indicates the branch of service with which the MTF is associated. This value is determined from the DMIS code list provided to DHSS by the EDC. It is missing when the Performing Facility information is missing. The possible values are: A=Army, C=Coast Guard, F=Air Force, M=Marine Corps, and N=Navy. This field is useful for limiting the observations included in an investigation. Often, the data available for use are limited by branch of service for the MTF or patient.

PERFORMING LOCATION FACILITY

The performing facility field in CHCS indicates the name of the MTF where the test was performed. This is a relatively standard text field. Problems are encountered if the text is entered incorrectly when the facility is registered in the system (i.e. misspellings). There is numerous facility names listed within the chemistry dataset but there are no values missing.

PERFORMING LOCATION WORK CENTER

The PERFORMING LOCATION WORK CENTER field indicates the work center within the laboratory that provided the service. This field is an unstructured text field with many possible values. Currently, there are 299 different facilities which conduct laboratory tests within the DOD. These locations are mapped according to the PERFORMING DMIS LOCATION.

RACE

There are six possible values for RACE: C=White, M=Asian or Pacific Islander, N=Black, R=American Indian or Alaskan Native, X=Other, and Z=Unknown. Of the Chemistry records, none are missing a value for RACE. However, 42.0% of the records are classified as Unknown. Those entries which have no reported RACE are labeled as "Unknown." The "Unknown" responses are assumed to be pre-populated, to eliminate blanks within the database. These records are distributed across approximately 200 DMIS locations, which indicate that the problem is not a data quality problem at one MTF or on a specific CHCS server. This limits the ability to use the data to look at diseases or conditions that disproportionably affect one race.

RECORD TYPE

The value "LCH" for RECORD TYPE identifies the chemistry dataset. All entries in this dataset have the value of LCH in this field. There are no missing data in this field.

REFERENCE RANGE

The REFERENCE RANGE is an alphanumeric variable which 1) shows a comparative numeric value indicating a result or lesser than the norm, 2) a character string (ex: "less than" or "greater than 8"), or 3) a positive or negative reference. These values are used in conjunction with the TEST RESULT to allow clinical interpretation of the results. The positive or negative readings have a character string entry in the TEST RESULT. Listings akin to "less than" have a titer result within the TEST RESULT, which are usually a derivative of 8. Other values show a

normal range or a normal titer level (1:8 or <8). There are different expressions of REFERENCE RANGE, and the majority of the entries are missing values.

REQUESTED DATE

The REQUESTED DATE is a date field formatted as YYYYMMDD, and includes one missing value. This field is not frequently used within the data analysis, as the ICD does not provide a detailed definition. All entries for this field have valid dates. The values of the REQUESTED DATE are between ORDER EFFECTIVE DATE and COLLECTION DATE.

REQUESTED TIME

This field represents the time component of the REQUESTED DATE formatted using a standard 24-hour clock. The timeframe is from 0000 to 2359, and all times are valid entries.

REQUESTING DMIS FACILITY NAME

This field is the text translation of the DMIS ID provided in the REQUESTING DMIS ID field. This field indicates the laboratory facility name that is requesting laboratory service to be completed. This allows for more accurate investigations when geographic information is used, because it is created using an official DOD DMIS list. This field is a translation of the Requesting Facility field in CHCS; therefore, it will be missing when that field is missing in the record.

REQUESTING DMIS ID

The REQUESTING DMIS ID is a four digit code assigned by the DOD to all units in all installations to uniquely identify them. The code allows grouping of MTFs based on geographic location, as well as to identify parent/child relationships between installations. Importing this field in character format can prevent the loss of leading zeros, which may produce complications when producing summary statistics.

Since this field is calculated based on the Requesting Facility field, all records missing a value for that field are missing a value for the REQUESTING DMIS ID field. Missing values are limited and seen at few specific MTFs. Other records include test entries from inactive sites (i.e. DMIS ID = 1007, ACH BAUMHOLDER).

REQUESTING FACILITY NAME

The REQUESTING FACILITY NAME is the field in CHCS that indicates the name of the MTF where the order originated, and is a relatively standard text field. Problems are encountered if the text is entered incorrectly when the facility is registered in the system (i.e. misspellings). The field allows tracking of orders from origin to where they are filled. There are 3 entries with missing values in chemistry.

REQUESTING FACILITY SERVICE

The REQUESTING FACILITY SERVICE field indicates the branch of service that the MTF is associated. This value is determined from the DMIS code list provided to DHSS by the EDC. It is missing from a record when the Requesting Facility information is missing. The possible

values are: A=Army, C=Coast Guard, F=Air Force, M=Marine Corps, N=Navy. Because this field is mapped to the REQUESTING DMIS FACILITY NAME and the REQUESTING DMIS ID, the REQUESTING FACILITY SERVICE is missing when the other two fields are blank. This field is useful for limiting the observations by the branch of service. This may be necessary for comparison, as data compared to the HL7 datasets are often service branch specific.

REQUESTING WORK CENTER NAME

The Requesting Work Center is the ward or clinic within the MTF that requests the laboratory test. This field is an unstructured text field with many possible values. Entries are labeled as DMIS ID number, clinic wards, service centers, and unknown/other MTF locations.

RESULT NOTES

The RESULT NOTES field is a character string which allows the laboratory technician to provide additional information about the result, a recommendation for additional testing, or the interpretation of the laboratory result. This field is either an automatic/drop down menu that is created when the TEST NAME is selected via CHCS or free-text. A frequency of the variable shows numerous variations of the same basic note. There are duplicates of long interpretations of a particular test by its DMIS location, which signifies syntax mapped out for the CHCS location and the test ordered. Other text entries show limited comments, such as the laboratory technician's initials, which genus type was found, or the address where the sample is tested.

The RESULT NOTES field is useful when the TEST RESULT field does not indicate the outcome of a test (i.e. when Test Result shows either a reading of 'SEE NOTES' or 'SEE COMMENTS'). The RESULT NOTES field is a hindrance due to the variability of the entries as well as having elongated text fields which limits the use of wildcards and searches. Nearly 83% of records are blank in this field, which indicates either a result's notes is not mapped to that particular test by the DMIS location or the laboratory technician did not manually enter information.

RESULT STATUS OBX

The RESULT STATUS OBX field is a character string which shows the status of the test performed. There are three entries which are used: P (Preliminary), F (Final), and C (Correction). These tests are used in a timely fashion, and always follow the order P, F, and then C. A test always has an F within a group of SET IDs of a test, but may also have a P or a C. Should a test have more than one RESULT STATUS OBX, it has the same SET ID, TEST NAME, and TEST ORDERED, but is on a separate entry line. An entry of "C" is entered when the record is amended due to operator error, wrong test ordered, or the test is performed for the wrong patient. An entry of "C" may also be entered if an error is identified in what was originally entered as the result.

Within the chemistry dataset, there are only F and C values present for RESULT STATUS OBX. The value of F is present for the majority populated within the dataset. There are no missing values within this field.

SENSITIVE RESULT FLAG

The SENSITIVE RESULT FLAG field remains unpopulated but is included in the final daily feeds.

SERVICE

The service field refers to the service branch of the sponsor. The value is determined from the first component of the PATCAT field and the values are the same. Therefore, there are an equal number of records that are missing the branch of service and PATCAT CODES. The highest proportion of records belonged to the Army, Navy/Marine Corps, and Air Force, respectively.

SET ID

The SET ID is a numeric field ranges from 1-9999. This field is automatically populated by CHCS; there are no missing values. The numbers show the logical order of arrival of OBX segment data within an HL7 message. There could be numerous types of tests and controls within the same ACCESSION NUMBER and SPECIMEN SOURCE, thus would have a different number associated to each test listed. Should an entry have a change in its resulting status (from pending to final, from final to corrected), the SET ID is the same for that test entry.

SPECIMEN SOURCE

The SPECIMEN SOURCE is a text field which describes the type of specimen tested. This field is useful to determine if the proper protocol is used for a laboratory test. For example, a SPECIMEN SOURCE with an entry of spine is not appropriate for a test for Orthomyxo influenza. The entries indicate tissue of the patient's body: nasal, naris, nose, pharynx, and throat. The laboratory sporadically does not differentiate between the SPECIMEN SOURCE and the specimen location (BODYSITE COLLECTION).

SPONSOR ID

The SPONSOR ID field corresponds to the SSN of the sponsor and is in a 9-digit format with no dashes. The SPONSOR ID is not sufficient to identify a unique patient, but may be used in conjunction with the FMP as a unique patient identifier. It is important to preserve the entire SSN when importing the data into any analysis program. The SPONSOR ID variable needs to be imported as a character field so that leading zeros are not dropped.

Not all Sponsor IDs are Social Security Administration SSNs. If the patient does not have a valid SSN, a pseudo SSN is created. The pseudo Sponsor ID begins with 800 or 900, followed by the date. If the number is already assigned to another patient, the primary three numbers will change to 801 or 901 consecutively depending on the number created with the same date.

Additionally, quality assurance testing is conducted in laboratories. Quality assurance procedures utilize SSN-like identifiers in the SPONSOR ID field. The Sponsor ID for these procedures may resemble a pseudo-SSN, arbitrary identifiers such as 777777777, or three consecutive zeros. These tests will have labels such as: Ztest, Quality Control, PSR, CAP, Non-human (NH,#), etc.

TEST NAME

The TEST NAME is a text field that shows which test is performed on the sample provided. This value is usually from a drop-down list of tests related to the TEST ORDERED variable. This field does not have missing values because TEST NAME is automated by the regional CHCS. The TEST NAME includes entries such as tests to be performed, quality controls, temperature, and even alerts for positive results. Quality control tests are within this field, and are noted via a ZZZ prior to the actual test name.

The variance between tests names suggests the fields are automated by a local CHCS host, not at the main location. A test procedure can be specific or general. A test name can refer to the procedure type, e.g. convalescent, cultures, or antibody testing. Therefore, a test name should match the type of test result present (a titer should be numeric, organism specific tests can reflect positive or negative or even sub-typing).

TEST ORDERED

TEST ORDERED identifies the requested observation, test, or panel. Each regional CHCS location has the autonomy to determine the criteria for each test ordered. Therefore, the TEST ORDERED field can have different grouping of tests per DMIS locations. The TEST ORDERED value is repeated among all records for tests associated with it according to the ORDER NUMBER. A provider can use a pull-down menu to determine the test(s) to be performed on a specimen. This shows all available tests per each test ordered.

TEST RESULT

The TEST RESULT is an alphanumeric field which shows either the pending information or the results of a test ordered. The TEST RESULT field can show positive or negative results, titer numbers, control values, dates, reorders, references to see comments, that a test was not performed due to inadequate results, or insufficient quality. This field is either automated via each regional CHCS location or entered manually. There are multiple variations, including misspellings and slang language (i.e. NOPERS). Many of these variations show the same result, such as Positive, POSITIVE, POS, and so forth. Currently, CHCS is in the process of regulating the regional CHCS locations to create one specific result for each TEST NAME outcome. This will limit the variation of TEST RESULT significantly.

There are 0.01% of the values missing for this variable. When the TEST RESULT is missing, it is recommended to consult the RESULT NOTES field, which is a free text field and may include the results of the test.

UNIT OF MEASURE

The UNIT OF MEASURE is a character field which shows the concentration of the TEST RESULT. Of the entries, 20.5% are blank. There are nine values listed, consisting of lower and upper case versions of L (Liter), M (minimum or mg/mL), R (reference range), T (Titer), and U (ug/mL). A percentage mark (%) makes up 0.4% of all entries. These values only compliment a numeric result located in the TEST RESULT, but not all numeric values within TEST RESULT have a UNIT OF MEASURE entry.

References

1. Center for Disease Control and Prevention (C.D.C). (2006). Influenza Symptoms and Laboratory Diagnosis Procedures. Retrieved from the World Wide Web on August 23, 2007 from: http://www.cdc.gov/flu/professionals/diagnosis/labprocedures.htm

2. Center for Disease Control and Prevention (2007). "Rapid Diagnostic Testing for Influenza: Information for Clinical Laboratory Directors (Fact Sheet) (C.D.C.)." Retrieved July 30, 2007, from http://www.cdc.gov/flu/professionals/diagnosis/rapidlab.htm.

3. Military Health Services. (2007). "DMIS ID Table Home." Retrieved July 31, 2007, from http://www.dmisid.com/cgi-dmis/default.

4. World Health Organization (WHO). (2005, July 2005). "WHO Recommendations on the Use of Rapid testing for Influenza Diagnosis." Retrieved August 1, 2006 from http://www.who.int/csr/disease/avian_influenza/guidelines/RapidTestInfluenza_web.pdf.

Definitions

Department of Defense (DOD)

A DOD beneficiary is a sponsor, dependent, or civilian seen within any military treatment facility.

Department of the Navy (DON)

DON may refer to *a)* records pertaining to sponsors and family members of either the Navy or Marine Corps seen within a military treatment facility, or *b)* records pertaining to sponsors and family members of any service seen within a Navy military treatment facility.

Defense Medical Information System (DMIS)

DMIS IDs are recognized within the DOD as the controlling standard for both medical and military facility identification and cost/workload classification. DMIS IDs are used throughout the MHS and worldwide in both healthcare and non-healthcare systems.

Abbreviations

AMA	American Medical Association
CAP	College of American Pathologists
CBC	Complete Blood Count
CDC	Centers for Disease Control and Surveillance
CHCS	Composite Healthcare System
CPT	Current Procedural Technology
DHSS	Defense Health Services System
DMIS	Defense Medical Information System
DOD	Department of Defense
DON	Department of the Navy
EDC	EpiData Center
EIA	Enzyme Immunoassay
FMP	Family Member Prefix
GEIS	Global Emerging Infections Surveillance and Response System
HL7	Health Level 7
ICD-9-CM	International Classification of Diseases, 9th Edition, Clinical Modification
MEPRS	Medical Expense Performance Reporting System
MHS	Military Health System
MTF	Military Treatment Facility
NMCPHC	Navy and Marine Corps Public Health Center
SSN	Social Security Number

Example project using Chemistry Laboratory data

Burden of Hepatitis A Virus (HAV) among DON Beneficiaries, CY 2011

Background

Hepatitis A is an acute liver disease caused by hepatitis A virus (HAV). It is transmitted primarily by ingestion of infected feces. In addition to common symptoms associated with gastrointestinal infections, those with HAV may also experience jaundice and fatigue. The guidelines for case ascertainment require a combination of both clinical and laboratory based criteria.

Purpose

The purpose of this project was to describe the burden of HAV among DON beneficiaries during CY 2011 using chemistry laboratory data and inpatient/outpatient medical records.

Methods

Standard Inpatient Data Record (SIDR) and Standard Ambulatory Data Record (SADR) databases were queried for all records of DON beneficiaries during CY2011 with International Classification of Disease 9th Revision (ICD-9) codes 070.0 (viral hepatitis a with hepatic coma) or 070.1 (viral hepatitis a without hepatic coma) in any diagnostic field. If both an inpatient and outpatient HAV encounter occurred on the same date, then the inpatient record was selected.

Health Level 7 (HL7) chemistry data that contained HAV-specific search terms with collection dates during CY2011 were extracted. Laboratory records were classified according to the test result field. If the record could not be classified by the test result field, the result notes were investigated. Records were classified as 'Positive', 'Negative', or 'Suspect'. Suspect cases included those with equivocal or borderline results. Quality assurance and laboratory interoperability tests were removed from the analysis.

An HAV case in this study was defined as an encounter with a positive or suspect laboratory result within +/- 14 days of an encounter date. Suspect laboratory results with a matching encounter record were considered cases. Encounter records with a negative laboratory test result or without any laboratory testing were not included in the analysis. Individuals with positive or suspect laboratory results without a matching encounter record were investigated further. All encounter records within +/- 14 days of the laboratory specimen collection date, regardless of diagnoses, were identified for these individuals. The purpose of this approach was to identify any encounter records with non-specific codes associated with HAV symptoms.

Results

There were 23 individuals who were diagnosed by medical encounter record with HAV during CY2011. Of these, 17 had a matching laboratory record, but only five individuals tested positive for HAV. The volume of laboratory tests for HAV was far greater than the number of encounter records; during CY2011, there were 14,035 tests ordered for 13,454 people. While the vast majority of ordered laboratory tests had negative results (n=13,957, >99%), there were 29 positive and 33 suspect laboratory results with no HAV-specific encounter records. When all encounter records that occurred within two weeks of positive or suspect laboratory tests were

reviewed for these individuals, 46 records for 11 individuals had at least one ICD-9 code related to the additional groups previously described. Review of all ICD-9 codes in these records resulted in no additional cases.

Conclusions

In CY2011, there were five cases of HAV identified through MTF encounters and laboratory testing. This was one fewer case than observed in CY2010. Overall, there was a low occurrence of infection, which was expected for two reasons. First, HAV is a vaccine-preventable disease and those at greatest risk are likely to be vaccinated. Second, the overall incidence in the US is relatively low, which would also result in lower transmission of HAV.

One of the major strengths in this study was the use of laboratory data to confirm HAV cases. Furthermore, laboratory-identified cases without encounters identified through the first query were reviewed on a case-by-case basis for non-specific diagnoses. The use of three reviewers to identify additional cases served as a method of validation and reduced potential for bias in the study.

Future Directions

Future studies for HAV case identification may involve exploring the Hepatitis A vaccination history among AD service members. This project may also be expanded to include other services in order to provide a comparison overview of HAV burden.